Gait Recognition Based on Shape and Motion Analysis of Silhouette Contours

Sruti Das Choudhury, Tardi Tjahjadi
School of Engineering, University of Warwick,
Gibbet Hill Road, Coventry, CV4 7AL, United Kingdom.
Email: t.tjahjadi@warwick.ac.uk

Abstract

This paper presents a three-phase gait recognition method that analyses the spatio-temporal shape and dynamic motion (STS-DM) characteristics of a human subject's silhouettes to identify the subject in the presence of most of the challenging factors that affect existing gait recognition systems. In phase 1, phase-weighted magnitude spectra of the Fourier descriptor of the silhouette contours at ten phases of a gait period are used to analyse the spatio-temporal changes of the subject's shape. A component-based Fourier descriptor based on anatomical studies of human body is used to achieve robustness against shape variations caused by all common types of small carrying conditions with folded hands, at the subject's back and in upright position. In phase 2, a full-body shape and motion analysis is performed by fitting ellipses to contour segments of ten phases of a gait period and using a histogram matching with Bhattacharyya distance of parameters of the ellipses as dissimilarity scores. In phase 3, dynamic time warping is used to analyse the angular rotation pattern of the subject's leading knee with a consideration of arm-swing over a gait period to achieve identification that is invariant to walking speed, limited clothing variations, hair style changes and shadows under feet. The match scores generated in the three phases are fused using weight-based score-level fusion for robust identification in the presence of missing and distorted frames, and occlusion in the scene. Experimental analyses on various publicly available data sets show that STS-DM outperforms several state-of-the-art gait recognition methods.

Keywords:
Gait, silhouette, Fourier descriptor, histogram matching, dynamic time warping, Krawtchouk moments.

1. Introduction

Numerous markerless gait recognition methods have demonstrated that gait has sufficient discriminatory power for identifying a human subject from a distance using low resolution video sequences without interfering with the subject's activity when physiological biometrics, e.g., face, fingerprint and iris are not clearly visible [1, 2, 3]. However, variations of the subject's clothes, footwear and hair style add complexity to gait recognition, and the subject's physical and mental conditions, e.g., leg injury, drunkenness and pregnancy, distort the walking pattern [2, 4]. Gait recognition is also affected by occlusions in the scene, variations in viewpoint and walking speed, shape distortions due to carrying conditions, shadows under feet and change in ground surface. Furthermore, gait characteristics change with ageing. Thus, a robust gait recognition method needs to analyse bio-mechanical gait characteristics via static and dynamic pose changes of gait as in [5, 6, 7].

Gait recognition methods can be classified into model-based and model-free. Model-based methods (e.g., [7, 8, 9, 10, 11, 12]) characterise a subject by a structural model and a motion model to mainly analyse dynamics of gait [2]. The structural model represents the subject by a stick figure, ellipsoidal fits or a volumetric model based on the proportions of the human body parts, and measures time-varying gait parameters, e.g., gait period, stance width and stride length for gait signatures. The motion model is used to analyse kinematical and dynamical motion parameters of the subject, e.g., rotation patterns of hip and thigh, joint angle trajectories and orientation change of limbs [2]. These methods can reliably deal with occlusions, and are invariant to scale changes, rotational effects and slight variations in viewpoint. However, they are characterised by complex search and mapping processes. The model-free methods (e.g., [4, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]) analyse the spatio-temporal shape and motion characteristics of a subject's silhouettes without assuming any explicit model of the subject's body. Although the inter-subject discriminability of these methods are high, they are susceptible to variations in viewpoints and the subject's attire.

A gait period, i.e., the time interval between successive heel strikes of the same limb, provides strong gait characteristics in terms of deformation of the subject's silhouette shape and motion pattern. The popular shape descriptors used to analyse static shape characteristics are Procrustes shape analysis (e.g., [17, 18, 23]) and Fourier descriptors (FDs) (e.g., [24, 25]). Although spatio-temporal deformation of the subject's shape in a gait sequence provides better discriminative power than its kinematics, inclusion of dynamical motion characteristics improves the identification rate. Thus, we introduce a gait recognition method STS-DM that combines the spatio-temporal shape (STS) features of a subject's silhouettes with the subject's dynamic motion (DM) characteristics over a gait period using both model-free and model-based approaches to achieve robustness against the maximum number of challenging factors of gait recognition when compared to state-of-the-art gait recognition methods, namely robustness against small carried items, walking speed variations, shadows under feet, limited variations in clothing, segmentation noise, changes in ground surface, missing body parts, self-occlusions and distorted or missing frames due to presence of occluding objects in the scene. STS-DM operates on the lateral (i.e., profile) view of a subject since this view contains most of the significant gait characteristics.

Most gait recognition methods do not consider the subject's arm-swing and the self-occlusion caused by it. Thus, STS-DM introduces a novel analysis of angular rotation pattern of leading knee (ARPoLK) of silhouette contours for subject identification in the presence of across-day variations, e.g., clothing, footwear, hair style and ground surface, with a consideration of the subject's arm-swing. STS-DM analyses the shape of the silhouette contours at ten phases of a gait period via their low-pass filtered FDs to only retain their global shape information. STS-DM uses ellipses fitted to body segments at ten phases of a gait period for full-body shape and motion analysis which is invariant to boundary shape distortions due to segmentation errors and missing or distorted body parts. Contour shape analysis at the ten phases that reveal most of the distinguishable shape characteristics also enables STS-DM to benefit from speed-invariant shape sequence processing with reduced processing time and achieve robustness against missing or distorted frames due to occlusions. Since the dynamic motion characteristics of gait manifest over a gait period more than in discrete phases, ARPoLK analysis is performed over a gait period.

The proposed STS-DM is thus motivated by the need for a gait recognition method that addresses a wide varieties of challenging factors that limit the success of gait as a behavioural biometrics to reliably identify a subject in practical situations. The novelties of STS-DM are: (a) it effectively combines static shape characteristics with the local and global dynamic gait characteristics to achieve robustness against the maximum number of challenging factors; (b) it analyses the subject's shape by FDs, and uses phase-weighted magnitude spectra (PWMS) to generate a match score; (c) it introduces an experimentally supported procedure for detecting carried items and a component-based FD analysis based on anatomical studies to achieve invariance to all common types of small carrying conditions, and this level of invariance has not been addressed before; (d) it introduces ARPoLK analysis which is invariant to self-occlusions of the limbs of a walking subject, and hence captures the local dynamic gait signature very efficiently; (e) the use of ARPoLK analysis enables STS-DM to implicitly address subject's arm-swing, and the use of dynamic time warping (DTW) to obtain a match score which is invariant to walking speed; (f) it analyses the full-body shape and motion characteristics based on ellipse-fitting to body segments and uses Bhattacharyya distance histogram matching (BDHM) to obtain a match score; (g) the match scores obtained by PWMS, DTW and BDHM are combined using weighted sum rule of score level fusion for robust identification; (h) the robustness of STS-DM against missing frames is demonstrated; and (i) STS-DM provides competitive identification rates with reduced computational complexity.

The rest of the paper is organized as follows. Section 2 discusses related work and Section 3 presents STS-DM. Experimental results are analysed in Section 4 and Section 5 concludes the paper.

2. Related Work

Various gait recognition methods analyse static shape and dynamic motion characteristics of gait sequences to address variations in viewpoint, walking speed, carrying condition and clothing, as well as other covariate factors, e.g., segmentation noise, occlusions, low resolution, changes in ground surface and shadows under feet. The gait recognition methods that address variations in viewpoint either depend on a) extraction of gait features which are invariant to change in view [26, 27, 28]; b) learning mapping or projection relationship between the gait characteristics of one view to another based on view transformation [29, 30, 31]; and c) construction of a 3D model of a subject from 2D images captured from different views using multiple calibrated cameras [32].

The method in [26] which employs a probabilistic sub-gait interpersonal model to analyse sub-gaits, i.e., different parts of a silhouette, uses Bayesian networks. In addition to variations in view, it is also robust to missing body parts. The method in [27] determines the motion of a subject's lower limb based on anatomical positions of hip, shin and ankle for view-invariant gait recognition using a viewpoint rectification approach. However, the ankle of a subject is most likely to be occluded by the presence of shadows under feet. Since it is impossible to estimate the positions of hip and shin in the case of a subject either wearing a skirt or a long coat, and carrying an item in upright position, the method is also not robust against variations in clothing and carrying conditions. The method in [30] uses joint subspace learning technique to learn a subject's prototype of different views, and represents the subject as a linear combination of these prototypes for view-invariant gait feature extraction. The method in [32] represents the 3D pose of a subject by using a tree structure of human skeleton, where the joints are denoted as the nodes of the tree. The gait of a subject is simulated by a stick model. The method combines static gait characteristics obtained by anthropometric measurements of different body parts with the dynamic gait characteristics obtained by analysing the joint angle trajectories of lower limbs for identifying a subject based on linear time normalisation technique. In addition to variation in view, the method is also robust to self-occlusions and change in ground surface.

A few methods address variations in walking speed effectively. The method in [33] computes shape variation-based frieze pattern (SVB frieze) of the difference frames obtained by subtracting the key frames, e.g., double support stance frames from the series of subsequent frames of a gait period. The method in [34] replaces the centroid-based shape configuration of traditional Procrustes shape analysis with high-order shape configuration to take into account of dynamic gait characteristics. The method introduces a differential composition model for increased inter subject discriminability and uses Procrustes distance for identifying a subject. The dynamics normalisation based gait recognition (DNGR) method in [13] normalises gait dynamics using population hidden Markov model whose states represent specific gait stances over a gait period, and gait recognition is achieved by estimating the distances between two normalised gait signatures in linear discriminant analysis space so as to maximise intra-class discrimination of subjects. DNGR uses an eigenstance reconstruction model to smooth silhouettes and achieves invariance to walking speed and changes in ground surface. The speed-invariant gait recognition method based on Procrustes shape analysis (SI-PSA) [35] uses Procrustes shape analysis based on high-order derivative shape configuration. The method using silhouette transformation based on walking speed (ST-WS) [36] separates static and dynamic features by fitting a human model and uses a factorization-based transformation model to transform the dynamic features from a reference speed to a target speed. The speed-invariant method in [37] uses the features extracted by Fisher discriminant analysis based cubic high-order local auto-correlation of the gait sequences to train a hidden Markov model.

Different approaches have been used to address variations in carrying condition. The method based on spatio-temporal motion characteristics, statistical and physical parameters (STM-SPP) [23] analyses the shape of a silhouette contour using Procrustes shape analysis at the double support phase and elliptic Fourier descriptors (EFDs) at ten phases of a gait period. A part-based EFD analysis is used to address shape distortion due to carrying conditions. The method in [9] uses models to obtain skeleton parameters by wavelet decomposition of a gait energy image (GEI) and invariant moments for combining anatomical and behavioural characteristics of gait. Thermal imaging is used to extract silhouettes that are invariant to carrying conditions and lighting variations. An iterative local curve embedding algorithm is used in [38] to extract double helical signatures.

A significant drop in recognition performance on the well-known public datasets, i.e., HumanID gait challenge dataset, MIT dataset, is reported in methods such as [15] when time covariate is encountered. However, there was no restriction on the subjects' clothing in any of these experiments. Therefore, the method in [39] analysed the effect of elapsed time on gait recognition in the absence of other covariates including clothing variations and concluded that gait successfully meets the criterion of a biometrics for reliably identifying a human subject at a distance over a considerable time interval. Based on experimental analysis, the method concluded that variation in clothing is the most challenging factor for model-free gait recognition methods.

The discriminability of a subject decreases due to shape distortions caused by clothing variations over different days. Therefore, the method in [40] uses an adaptive mechanism for combining part-based features to achieve robustness against clothing variations. The method in [19] uses radial integration transform, circular integration transform and weighted Krawtchouk with genetic algorithm (RCK-G). It assigns depth information captured by a calibrated stereo camera to binary silhouettes using 3-dimensional (3D) radial silhouette distribution transform and 3D geodesic silhouette distribution transform. Genetic algorithm fuses the 2-dimensional (2D) and 3D features extracted by radial integration transform, circular integration transform and weighted Krawtchouk moments. RCK-G is robust to very limited clothing variations, but not insensitive to carrying conditions.

The methods in [4, 10, 11, 16, 20, 21, 22, 41, 42] achieve limited invariance to a few covariate factors for improved identification rate. The method in [10] uses appearance and dynamic traits of gait by analysing parameters of ellipses fitted to seven regions of a subject's silhouette, i.e., centroid, aspect ratio and elongation along with the subject's height for identification which is invariant to limited clothing variations and segmentation imperfections. The method in [11] uses a full-body layered deformable model to analyse the widths, lengths, positions and orientations of ten parts of the subject for manually labelled and automatically extracted silhouettes. The method addresses self-occlusions, and the incorporation of upper body dynamics in addition to limbs enables the method to achieve robust identification in the presence of variations in footwear, clothing, ground surface and time.

The performance of a gait recognition method deteriorates if: a) the captured gait sequences are of very low resolution either due to the low resolution of the camera or large distance between the subject and the camera; and b) the inter-subject discriminative information is reduced due to the projection of gait sequences onto nonoptimal low-dimensional subspace in order to reduce the dimensions of the feature space. The method in [43] thus uses superresolution with manifold sampling and backprojection to transform low resolution gait sequences into high resolution, and incorporates nonparametric multilinear tensor-based dimensionality reduction technique for improved identification rate.

The method in [4] captures spatio-temporal motion information of a gait period in a single GEI and the method in [41] captures temporal information of a gait sequence in a single multichannel chrono gait image (CGI). The methods GEI [4] and CGI [41] manually compute synthetic gait templates by employing a cutting and fitting scheme based on anthropometry to take into account distortions of lower body part due to carrying a briefcase, and variations in ground surface, clothes and footwear, but not distortions of upper body-part due to variations in clothing and carrying condition. A GEI is noise-resilient, and its use enables a method to be computationally less expensive in terms of time and space. The method CGI uses a gait period detection technique that is robust to shadows under feet and carrying a briefcase. The method in [20] uses a set of local augmented Gabor features extracted from different scales and orientations to characterise a GEI, and uses a new patch distribution feature for subject identification. A locality-constrained group sparse representation is introduced to address the presence of different covariate factors, e.g., change in ground surface and carrying a briefcase. The method in [42] enhances the dynamic information content of GEI by computing gait entropy image for identifying a subject in unconstrained environment with limited variations in the covariates over different days, but performs poorly in the presence of changes in viewpoint.

The method in [22] computes gait flow image (GFI) from binary silhouettes using optical flow field for identifying a subject. The method based on general tensor discriminant analysis and Gabor features (GTDA-GF) [21] uses the following image representations for feature extraction: sum of Gabor filter responses over directions; sum of Gabor filter responses over scales; and sum of Gabor filter responses over scales and directions. The methods GFI [22] and GTDA-GF [21] outperform the method GEI for the cases of carrying a briefcase, variations in viewpoint and footwear of the HumanID gait challenge data set. The method in [44] considers gait sequences as a third-order tensor to introduce a gait representation called EigenTensorGait followed by application of linear discriminant analysis for gait recognition using multilinear principal component analysis based tensor object recognition framework. The method is robust against limited variations in viewpoint and footwear of the subjects. The method in [6] uses prediction-based hierarchical active shape model (ASM) and Kalman filtering to achieve invariance to illumination variations, shadows and considerable occlusions.

While the trend of the state-of-the-art gait recognition methods is to address only one or a few covariate factors, STS-DM first attempts to identify a subject in presence of a wide range of challenging factors with low computational complexity for practical deployment. It fuses the local and global gait characteristics obtained by analysing static shape and dynamic motion of silhouette contours to address the maximum number of covariate factors so as to achieve combined invariance to carrying conditions, walking speed, shadows under feet, limited variations in clothing, segmentation noise, changes in ground surface, missing body parts and occlusions. Like the method in [24], STS-DM also characterises a subject's shape using FDs but introduces a novel component-based FD analysis to achieve invariance to all common types of small carrying conditions. STM-SPP [23] and the method in [18] only analyse the static shape characteristics of a subject, but STS-DM analyses the local and global dynamic motion characteristics with a consideration of arm-swing in addition to static shape characteristics to achieve robustness against more across-day gait variations. Since the upper body dynamics also play a significant role in gait recognition [11], similar to the method in [10] STS-DM uses the orientation angle, aspect ratio, area and eccentricity of the ellipses fitted to five segments of a subject's silhouette contour to analyse the shape and motion characteristics of the entire body in addition to local dynamic gait analysis by ARPoLK. The advantage of contour-based ellipse-fitting over region-based ellipse-fitting as in [10] is low computational complexity. Following the attempt in [5] which combines static and dy-

Table 1: Acronyms used in the paper.

Acronym	Description
ARPoLK	angular rotation pattern of leading knee
ASM	active shape model based gait recognition in [6]
BDHM	Bhattacharyya Distance Histogram Matching
CCR	Correct Classification Rate
CASIA	Silhouette analysis-based gait recognition in [18]
CMU	silhouette-based human identification from body shape and gait in [55]
CMU1	gait recognition based on shape estimation in [57]
CGI	chrono gait image based gait recognition in [41]
DNGR	dynamics normalisation based gait recognition method in [13]
DTW	Dynamic Time Warping
EFD, FD	Elliptic Fourier Descriptor, Fourier Descriptor
GTDA-GF	general tensor discriminant analysis and Gabor features based gait recognition in [21]
GEI	gait energy image based gait recognition in [4]
GFI	gait flow image based gait recognition in [22]
MMFA	gait recognition using matrix-based marginal Fisher analysis in [14]
PWMS	Phase Weighted Magnitude Spectra
Rf-ROI	reference Region-of-Interest
RCK-G	radial integration transform, circular integration transform and weighted Krawtchouk moments with genetic algorithm based gait recognition in [19]
STM-SPP	spatio-temporal motion characteristics, statistical and physical parameters based method in [23]
SI-PSA	speed-invariant gait recognition method based on Procrustes shape analysis in [35]
ST-WS	silhouette transformation based walking speed invariant gait recognition in [36]
SSM	shape sequence matching based method in [17]
SVB frieze	gait recognition using shape variation-based frieze pattern in [33]
SSP	image self-similarity plot based gait recognition in [16]
Tr-ROI	target Region-of-Interest

namic gait signatures, STS-DM uses weight-based sum rule of score-level fusion to fuse the match scores obtained by different classifiers for subject identification. To demonstrate the efficacy of STS-DM in terms of robustness against most of the challenging factors that affect existing gait recognition systems, it is compared with several related state-of-the-art gait recognition methods which are referred to by their acronym for brevity. Table 1 lists the acronyms of these methods as well as the other acronyms used in the paper.

3. STS-DM

STS-DM comprises three modules as shown in Fig. 1. Module 1 extracts and postprocesses silhouette contours. Module 2 extracts gait features in three phases. Phase 1 analyses spatio-temporal changes of a subject's shape based on PWMS of FDs of the silhouette contours to generate a match score. Phase 2 performs full-body shape and motion analysis, and compares probe and gallery gait signatures by BDHM. Phase 3 uses DTW to measure similarity between ARPoLKs of the probe and gallery subjects. The match scores generated in three phases are fused using weight-based score-level fusion in module 3 for subject identification.

3.1. Module 1: Extract and postprocess silhouette contours

The performance of a contour-based method can be substantially enhanced if the contours are extracted from high quality silhouettes, i.e., silhouettes without shadows, missing body parts and parts of the background [45].

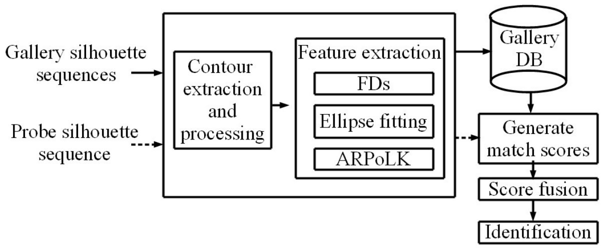


Figure 1: Overview of STS-DM.

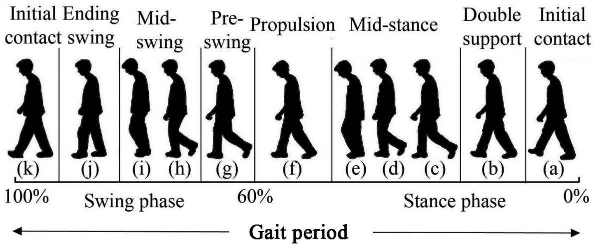


Figure 2: Ten phases of a gait period (a)-(k) of a subject from CMU MoBo data set: stance phase (a)-(f); and swing phase (g)-(j).

Thus, the silhouettes from the data sets used to evaluate the performance of STS-DM are improved using eigenstance reconstruction model [13, 45]. The silhouette is then subjected to vertices traversal algorithm based on connectivity [46] to extract its extreme outer boundary, i.e., contour. To remove camera depth variations, the image is cropped according to the perimeter of the bounding rectangle enclosing the contour and resized to a fixed height while retaining the aspect ratio (i.e., ratio of silhouette width to its height) using bilinear interpolation. The retainment of aspect ratio ensures the maintenance of the proportional relationship between the width and height of a silhouette to preserve actual silhouette shape characteristics, which is a very important factor in shape based subject classification. The resized contour is then copied to a destination image of fixed size by coinciding its centre-of-mass with the centre of the destination image to make it translation invariant.

3.2. Phase 1 of Module 2: Analyse shape using FDs

A gait period begins with the heel strike of either foot and ends with the subsequent heel strike of the same foot. It consists of two steps, where a step is the time period between successive heel strikes of opposite feet. In a gait period, each foot transits between two phases: a stance phase and a swing phase, where when one foot is in stance phase (i.e., in contact with the ground) the other foot is in swing phase, as illustrated in Fig. 2. The stance phase begins with initial contact of heel of the foot making a forward movement (i.e., the forward foot) with the ground and ends with the toe lifting of the same foot from the ground. It consists of: initial contact when heel of the forward foot touches the ground; double support stance when both feet are almost flat on the ground and farthest from each other; mid-stance when the forward foot is initially positioned flat on the ground, carrying the body weight; and propulsion which begins with the heel lifting of the foot until prior to its toe off the ground indicating the termination of stance phase and start of swing phase.

During the swing phase, the foot does not remain in contact with the ground and the phase comprises: pre-swing which begins with heel of the forward foot off the ground and continues until maximum knee flexion; mid-swing, i.e., from maximum knee flexion to when the tibia is vertical to the ground; and ending swing, i.e., from vertical position of the tibia to just prior the forward foot makes initial contact with the ground. Similar to the method in [15], a gait period is determined by the number of frames between two frames of a gait sequence with the most foreground pixels enclosed in the region bounded by bottom of the bounding rectangle and the anatomical position of just before the subject's hand measured from the bottom (i.e., $0.377H$ where H is height of the bounding rectangle) because

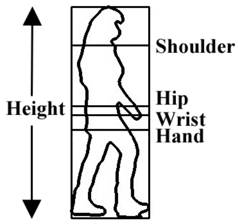


Figure 3: Anatomical positions of shoulder, hip, wrist and hand as a fraction of the subject's height are denoted by horizontal lines on a lateral-view of a walking subject's contour.

this foreground region, i.e., the bottom segment of the bounding rectangle is not distorted by self-occlusions due to arm-swing (see Fig. 3). The anatomical positions are determined when the subject is standing erect and at rest, with feet together and arms at the side, and the head and the palms of the hands facing forward. In Fig. 3, horizontal lines are used to denote the anatomical positions of shoulder, hip, wrist and hand as the fractions of a subject's height, i.e., $0.818H$, $0.530H$, $0.485H$ and $0.337H$, respectively, measured from the bottom of the bounding rectangle [47]. Note that these positions, which are based on anthropometry, might slightly deviate from the actual positions of the shoulder, hip, wrist and hand of a subject especially when the subject is performing an activity, e.g., walking as illustrated in Fig. 3.

The Krawtchouk moments of order $(n + m)$ of a $N \times M$ silhouette with intensity function $f(x, y)$ are computed using the sets of weighted Krawtchouk polynomials $\overline{K}_n(x; p, N)$ and $\overline{K}_m(x; p, M)$ as [19, 48]

$$Q_{nm} = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \overline{K}_n(x; p1, N-1).\overline{K}_m(y; p2, M-1).f(x, y), \tag{1}$$

where $n = 0, 1, ..., N$ and $m = 0, 1, 2, ..., M$. The set of weighted Krawtchouk polynomials, i.e., $\overline{K}_n(x; p, N)$ is defined as

$$\bar{K_n}(x; p, N) = K_n(x; p, N) \sqrt{\frac{w(x; p, N)}{\rho(n; p, N)}}, \text{where} \quad p \in (0, 1), \tag{2}$$

and

$$\rho(n; p, N) = (-1)^n \left(\frac{1-p}{p}\right)^n \frac{n!}{(-N)_n}. \tag{3}$$

Krawtchouk moments have better image reconstruction capability than the Zernike and Hu moments in both noisy and noise-free conditions, and the orthogonal property of weighted Krawtchouk moments ensures the minimal information redundancy [19, 48]. The scale and rotation dependency of Krawtchouk transform do not affect the extracted features as STS-DM considers only lateral views of silhouettes to achieve rotation invariance, and the silhouettes are pre-scaled and centre-aligned to achieve scale invariance. The Krawtchouk moments are also useful when dealing with partially distorted frames of a gait period, as they have the ability to extract local features from any Region-of-Interest (ROI) of an image by varying the parameters N and M.

The silhouettes of the ten phases in Fig. 2(a)-(j) are manually extracted. The bottom segment of the bounding rectangle is set as the reference Region-of-Interest (Rf-ROI) and the same silhouette segments of all frames of a subject's gait period are each referred to as a target Region-of-Interest (Tr-ROI). Unlike STM-SPP [23] which uses contour matching based on Hu moments for the detection of ten phases, STS-DM computes weighted Krawtchouk moments of each of the Rf-ROIs and Tr-ROIs using Eq.(1) by suitably choosing the values of N (say, c) and M (say d) (such that they respectively denote the width and height of the bottom segment of the bounding rectangle) of order (c+d) using $p = 0.5$.

To obtain the ten phases of a gait period of any gait sequence automatically, the Rf-ROIs are compared with the target Region-of-Interest (Tr-ROI)) using silhouette comparison based on weighted Krawtchouk moments to obtain similarity scores [46]

$$S_{score} = \left[(\text{Rf-ROI}_{k_{nm}} - \text{Tr-ROI}_{k_{nm}})^2 \right]^{\frac{1}{2}}, \tag{4}$$

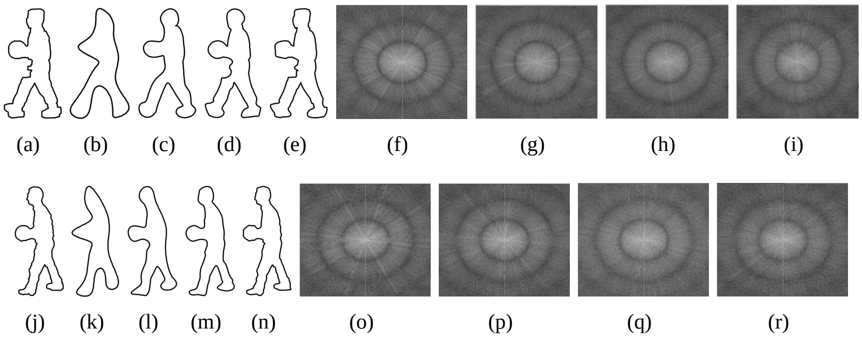


Figure 4: Reconstruction of contours using different number of FDs for subject 1(row 1) and subject 2 (row 2) from CMU MoBo data set: (a) and (j) Original contours with 2^8 points. Reconstructed contours using: (b) and (k) 2^4 FDs; (c) and (l) 2^5 FDs; (d) and (m) 2^6 FDs; (e) and (n) 2^7 FDs. Magnitude spectra of the contours with: (f) and (o) 2^4 FDs; (g) and (p) 2^5 FDs; (h) and (q) 2^6 FDs; (i) and (r) 2^7 FDs.

where Rf-ROI$_{k_{nm}}$ and Tr-ROI$_{k_{nm}}$ respectively denote the $(c+d)$ order weighted Krawtchouk moments of the Rf-ROI and Tr-ROI. The frame whose Tr-ROI results in the lowest S_{score} with the corresponding Rf-ROI is extracted as one of the ten phases, and the process continues by comparing the next Rf-ROI with the remaining Tr-ROIs until all ten phases are obtained.

The discrete Fourier transform of a contour results in a set of complex numbers, i.e., FDs which represent the shape of the contour in the frequency domain. FDs can be used to reconstruct the shape of the contour and are thus useful boundary shape descriptors for object recognition. Since the low-frequency (i.e., low-order) FDs contain global shape characteristics and the higher frequency (i.e., higher order) FDs increasingly contain finer shape details, a subset of FDs substantiates the discrimination between different shapes. Hence, we characterise a subject's contour using FDs to take into account of spatio-temporal change in the subject's shape over a gait period.

The silhouette contour points are traversed anticlockwise and each point with coordinates (x, y) is represented by a complex number $c(t) = x(t) + jy(t)$, where $t = 0,1,2,...,T-1$ and T is the number of contour points. The FDs are

$$a(u) = \frac{1}{T}\sum_{t=0}^{T-1} c(t)e^{-i2\pi ux/T}, \quad \text{for} \quad u = 0, 1, 2, ..., T-1, \tag{5}$$

where u is frequency variable. The original contour is restored by its inverse discrete Fourier transform, i.e.,

$$c(t) = \sum_{u=0}^{T-1} a(u)e^{i2\pi ux/T}, \quad \text{for} \quad t = 0, 1, 2, ..., T-1. \tag{6}$$

To ensure that all ten contours of a gait period are represented by a similar set of equal number of points, each contour is approximated by $T = 2^8$, i.e., 256 points using interpolation based on point correspondence analysis [24].

The magnitude and phase of FDs are respectively

$$|a| = [R_a^2(u) + I_a^2(u)]^2 \text{ and } \phi(a) = \tan^{-1}\left[\frac{I_a(u)}{R_a(u)}\right], \tag{7}$$

where $R_a(u)$ and $I_a(u)$ are the real and imaginary components of $a(u)$, respectively. The dynamic range of the magnitude spectrum is compressed using log operation and the resulting spectrum is translated to the centre of the Fourier plane to enhance its display in Fig. 4(e)-(g).

Fig. 4(b)-(e) respectively show the reconstruction of contours using 2^4, 2^5, 2^6 and 2^7 FDs of subject 1's original contour in Fig. 4(a), and Fig. 4(k)-(n) respectively show the reconstruction of contours using 2^4, 2^5, 2^6 and 2^7 FDs of subject 2's original contour in Fig. 4(j). Note that the use of just a few low-frequency FDs, e.g., 2^4 FDs results in very similar contours without any inter-subject discriminatory shape characteristics. However, as the number of

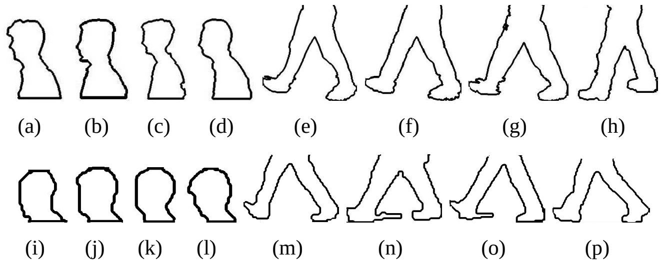


Figure 5: Row 1: (a)-(d) upper segments, and (e)-(h) lower segments of four subjects from CMU MoBo data set to exclude the ball. Row 2: (i)-(l) upper segments, and (m)-(p) lower segments of four subjects from CASIA gait data set B to exclude the backpack.

FDs increases, the contour gradually regains its original shape characteristics, and hence discriminability between different subjects also increases. Since global shape information is preserved in the low-frequency FDs, the use of the first half of the FDs for reconstruction results in the reconstructed contour which is almost the same as the original. We thus use an ideal lowpass filter to retain the first $T/2$, i.e., 2^7 FDs as they contain adequate subject-specific shape characteristics, while removing the higher-frequency FDs which contain the finer shape details. Their removal reduces flickering noise at the contour and smoothes contours from clothing curvatures. It also reduces the number of contour points to process.

The magnitude spectrum is multiplied by the corresponding phase to generate PWMS, the first gait signature. With initial contact (Fig. 2(a)) as the first phase of a gait period, PWMS provides the greatest variability between subjects, as it conveys additional information about temporal deformation of the sequence of shapes together with its frequency contents [49]. PWMS are represented as a $o \times k$ matrix, where o represent the ten phases and $k = T/2$, i.e., 128. Let $\mathbf{A}$ and $\mathbf{B}$ be two such matrices for a gallery and a probe gait sequences, respectively. The dissimilarity score between them is

$$\mathrm{d}_{\mathrm{PWMS}} = \frac{\sum_{i=1}^{o}\sum_{j=1}^{k}(\mathbf{A}_{i,j} - \mathbf{B}_{i,j})^2}{\sum_{j=1}^{k}(\sum_{i=1}^{o}(\mathbf{A}_{i,j} - \mathrm{mean}(\mathbf{A}_j))^2)}, \tag{8}$$

where $\mathbf{A}_j$ is jth column vector of $\mathbf{A}$, and mean(.) computes the average of the column vectors of $\mathbf{A}$. The range of $\mathrm{d}_{\mathrm{PWMS}}$ is [0,1], the smaller the value the more similar are the two shapes. We obtain dissimilarity scores by comparing the sequence of a probe subject with each sequence of the gallery subjects for a gait period, and the average dissimilarity score is used for classification.

When either a gallery or a probe subject carries an item, certain parts of the silhouette are altered and the discriminability of the gait recognition algorithm decreases with respect to the affected parts of the silhouette. Therefore, we introduce a component-based FD analysis based on anatomical studies of human body to make STS-DM robust to carrying conditions. If a subject carries a small item with folded arms or a backpack, the shape of the silhouette above the anatomical position of shoulder, i.e., $0.182H$, and below the anatomical position of wrist, i.e., $0.515H$, measured from top of the bounding rectangle are not affected. The validity of this assumption is experimentally verified on the CMU MoBo data set for the subjects carrying a ball with folded arms and on CASIA gait data set B [50] for the subjects carrying a backpack. Thus, carrying a backpack or a small item with folded arms by a subject can be detected by analysing the difference in the number of contour points enclosed in the region bounded by the top of bounding rectangle and the anatomical position of wrist. It is experimentally shown that the upper segment of the silhouette enclosed between $0.225H$ from top of the bounding rectangle, and the lower segment enclosed between $0.500H$ and bottom of the bounding rectangle are not affected by the presence of a ball for all twenty four subjects walking with ball in the CMU MoBo data set (Fig. 5(a)-(h)). Thus for component-based FD analysis, the shape of a silhouette is segmented into an upper segment spanning from top of the bounding rectangle up to the shoulder, and a lower segment spanning from the anatomical position of wrist to bottom of the bounding rectangle. The average of the dissimilarity scores of these components is used for subject classification.

The shape of a silhouette above the wrist is not altered when the subject's hand carries a small bag. According to anthropometry, the position of the wrist is estimated to be $0.485H$ [47] measured from bottom of the bounding rectangle. Thus, an analysis of silhouettes enclosed in the region (1-0.485H), i.e., 0.515H, measured from the top of the bounding rectangle using FDs, remove the shape variations due to carrying briefcase among the gallery and probe sequences. Analysing the leg region, i.e., between aH and bH (where a = 0.750 and b =0.9375 [51]), removes the effect of shape distortion due to the presence of a briefcase and shadows under feet. Thus, STS-DM detects the presence of a small bag by examining the difference in the number of contour points enclosed in the region between 0.515H and 0.750H. A substantial increase in the number of contour points (e.g., for USF data set an increase of at least twenty points of non approximated contour and for the same phase of a gait period between gallery and probe sequences) confirms the presence of a briefcase.

It is to be noted that the method in [38] analyses symmetry changes in double helical signatures at the limb region to take into account of shape distortion due to a specific carrying condition, e.g., a briefcase by an upright subject. The use of synthetic gait templates in GEI and CGI, manually computed by simulating distortions in the lower body part of the silhouettes, enables these methods to achieve invariance to the distortions in the lower part of the body, but not in the upper part, e.g., due to carrying conditions with folded arms. The component-based FD analysis in STS-DM and part-based EFD analysis in STM-SPP [23] both rely on anatomical studies of human body to achieve invariance to carrying conditions. However, STM-SPP takes into account of carrying conditions either with folded arms or in upright position, but does not consider subjects carrying a backpack. Hence, STS-DM provides the most in-depth analysis of invariance to carrying condition by taking into account of all common types of small items carried by a subject on the back, with folded ams and also in upright position.

3.3. *Phase 2 of Module 2: Analyse full-body shape and motion*

Undoubtedly, lower body dynamics capture the most distinguishable gait characteristics, but consideration of shape and motion characteristics of upper body enhances it. Therefore, the shape and motion characteristics of the full-body contour is analysed by parts at ten phases of a gait period for extracting global gait signatures. To take into account of change in appearance of different parts of a contour due to walking, the contour is divided into four regions with each region fitted with an ellipse. An ellipse is preferred to a circle and a rectangle as it has more useful parameters to describe shape characteristics (i.e., aspect ratio, area and eccentricity) and motion characteristics (i.e., orientation angle, the angle of the semi-major axis of the ellipse measured anti-clockwise from the positive horizontal axis). Also, ellipse fitting approach is robust to limited distortions at the contour due to poor segmentation, and enables STS-DM to take into account of subject-specific characteristics, i.e, fat vs slim and long hair vs bald.

The height H of the bounding rectangle enclosing a silhouette contour is used as the subject's height. Following anatomical studies of the human body [47], the vertical positions of shoulder, hip and knee measured from the bottom of the bounding rectangle are estimated respectively to be 0.818H, 0.530H and 0.285H. The bounding rectangle is then subdivided into the following four regions (as shown in Fig. 6(a)) by drawing horizontal lines at the anatomical positions of shoulder, hip and knee joints: uppermost region enclosing head and neck; region enclosed between shoulder and hip; region enclosed between hip and knee; and bottommost region enclosing the legs. The bottommost region is subdivided by a vertical line into two regions, each enclosing one leg. The vertical line if extended, passes through the centre-of-mass of the contour. The process is illustrated in Fig. 6(a)-(b), where centre-of-mass is abbreviated as COM. The 2D Cartesian moment of order u and v of a contour $I(x, y)$ is

$$m_{u,v} = \sum_{i=1}^{T} I(x, y)x^u y^v. \tag{9}$$

The centre-of-mass of the silhouette contour, (x_c, y_c), is given by $x_c = \frac{m_{10}}{m_{00}}$ and $y_c = \frac{m_{01}}{m_{00}}$ [52].

The contour points enclosing each of the five regions representing a body part are best fitted by an ellipse using a non-linear least squares technique as illustrated in Fig. 6(c)-(d). The following four parameters of each of the fifty ellipses for the ten phases form the gait signature: aspect ratio; area; eccentricity; and orientation angle.

We compute twenty 1D histograms, each representing the distribution of one parameter of one ellipse (i.e., one segmented region) for ten phases of a gait period. The histograms are normalised to [0,1] as shown in Fig. 7. The normalised histograms of the probe gait sequences (Hist-p$_n$, n=1,..., 20) are compared with the corresponding histograms

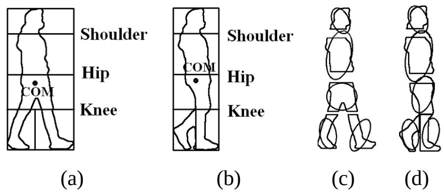


Figure 6: (a)-(b) Partitioning a subject's contour into five segments (COM denotes centre-of-mass); (c)-(d) ellipses fitted to each of the five segments.

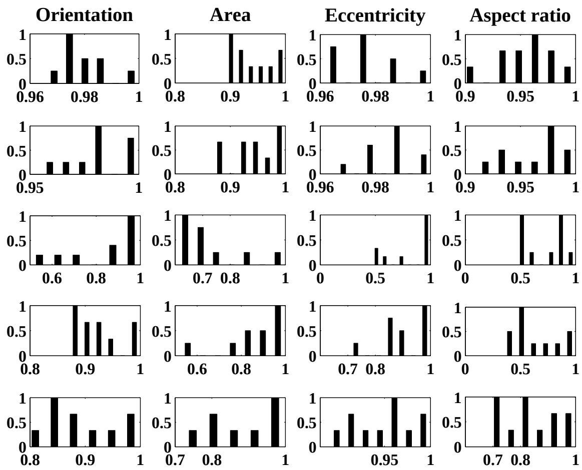


Figure 7: The histogram matrix with each column representing histograms of ellipse parameters (orientation angle, area, eccentricity and aspect ratio), each for ten phases, corresponding respectively from top to bottom to the following regions: head-shoulder, shoulder-hip, hip-knee, right leg and left leg.

of the gallery sequences (Hist-g$_n$) using Bhattacharyya distance metric to obtain the dissimilarity score [46]

$$d_n(\text{Hist-p}_n, \text{Hist-g}_n) = \left(1 - \sum_{i}^{B} \frac{\sqrt{\text{Hist-p}_n(i).\text{Hist-g}_n(i)}}{\sqrt{\sum_i \text{Hist-p}_n(i).\ \sum_i \text{Hist-g}_n(i)}}\right)^{\frac{1}{2}}, \tag{10}$$

where B is the number of bins in each histogram. The second gait signature is the average dissimilarity score

$$\mathrm{d_{BDHM}} = \frac{1}{20}\sum_{n=1}^{20} d_n. \tag{11}$$

The range of $\mathrm{d_{BDHM}}$ is [0,1], and the low values of $\mathrm{d_{BDHM}}$ indicate good matches. Hence, ideally a probe subject would result in the lowest $\mathrm{d_{BDHM}}$ for its correct match in the gallery.

3.4. Phase 3 of Module 2: Analyse ARPoLK

We analyse ARPoLK to take into account of a subject's lower body dynamics which is robust to the problems associated with self-occlusions. Points h and k in Fig. 8 respectively correspond to where horizontal lines drawn across the bounding rectangle at heights of $0.530H$ (denoting hip position) and $0.285H$ (denoting knee position), measured from bottom of the bounding rectangle, meet with the contour facing the direction of walking. The angle of the leading knee θ_{LK} (i.e., q in Fig. 8) is subtended by the horizontal line through h and the line joining h and k. Measuring this angle over a gait period gives ARPoLK. The main motivation in using ARPoLK analysis is its ability to capture lower body gait dynamics which remain unaffected by self-occlusion, i.e., occlusion of one knee

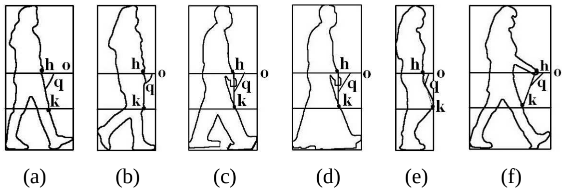


Figure 8: Illustration of ARPoLK: (a)-(b) subject 1 at two different phases of a gait period of two different sequences with different hair style and shoe type on different days; (c)-(d) subject 2 with and without shadows under feet in two different gait sequences; and (e)-(f) subject 3 with higher arm-swing.

by another. ARPoLK analysis always takes into account of only the leading knee, i.e., the knee which is at the front in the direction of walking, and does not consider the other knee which might be occluded by the leading knee at times during walking. Fig. 8(b) and (e) demonstrate the successful ARPoLK analysis in the case of an occluded knee by the leading knee. ARPoLK analysis is also invariant to across-day gait variations that affect the subject's shape above the hip and below the knee, e.g., change of hair style and shadows under feet, and takes into account of limited changes in clothing style, such as pant vs skirt or shorts across different days, but not subtle style change like tight vs loose clothing. Fig. 8(a)-(b) illustrate that ARPoLK analysis remains unaffected by change of hair style and footwear including the use of high heels, while Fig. 8(c)-(d) illustrate similar values of θ_{LK} for the same subject's silhouette with and without shadows under feet. ARPoLK analysis is also invariant to carrying of small items with folded arms as long as the subject's hip is not occluded.

Most but not all gait recognition methods that consider dynamic gait characteristics only focus on the motion of lower limb region, but ignore arm-swing although it is an unavoidable and integral part of gait. In normal walking, the contralateral arm is automatically swung forward with the swinging lower limb at a rate proportional to the walking speed, and different subjects have varying arm swings. Therefore, arm-swing is integrally related to the motion of lower limbs, and contributes to inter-subject discrimination. The method in [7] uses linear Hough transform to model arm-swing as a pendulum motion. But arm-swing can be arguably modelled as a pendulum motion because it is considerably influenced by neuromuscular forces. Although the method uses hip and shin angles to constrain the Hough space, it models the limb motion and arm-swing separately, and therefore, does not consider the integral relationship between them. The method in [19] indirectly addresses the integral relationship between arm-swing and motion of lower limb in gait by holistic image analysis. STS-DM also considers this integral relationship by analysing the lower limb motion in conjunction with the arm-swing using ARPoLK analysis. ARPoLK analysis implicitly addresses the effects of arm-swing as evident from Fig. 8(e)-(f). However, the effect of arm-swing is not considered in ARPoLK analysis if the hands of a subject are engaged due to carrying conditions either with folded arms or in upright position. Hence, the subject identification performance of ARPoLK analysis is significantly affected if a subject has a higher arm-swing as in Fig. 8(f) in a gallery sequence but his/her hands are engaged by carrying conditions in a probe sequence, and vice versa.

Fig. 9 shows the discrete signals obtained by ARPoLK analysis of two different gait sequences corresponding to subject 1 (Fig. 8(a)-(b)) and subject 2 (Fig. 8(c)-(d)). The discrete signals are formed by plotting the different values of θ_{LK} for a gait period against N equally spaced monotonically increasing values, where N is the number of frames in a gait period. The signals are normalised in the range [0,1] by dividing each θ_{LK} with the maximum value of θ_{LK} to remove spatial scale variations for different subjects for uniform comparison. It is evident from Fig. 9 that discrete signals representing ARPoLK of two different gait sequences of the same subject resemble each other, while different subjects have quite dissimilar signals. Thus it is verified that ARPoLK analysis has a very good inter-subject discriminability in the presence of across-day gait variations and shadows under feet.

Different subjects have different walking speeds which result in varying number of frames in their gait period. Depending on the subject's state of mind, the number of frames in a gait period of the same subject also varies due to the walking speed variations in different situations. Therefore, we use DTW to account for such variations in classifying a probe subject based on its similarity of ARPoLK with that of a gallery subject over a gait period. DTW uses dynamic programming to compute a warping function that optimally aligns two time-dependent sequences of varying lengths for measuring similarity. Let $a_g = \theta_{LKg1}, \theta_{LKg2}, ..., \theta_{LKgm}$ and $a_p = \theta_{LKp1}, \theta_{LKp2}, ..., \theta_{LKpn}$ be the

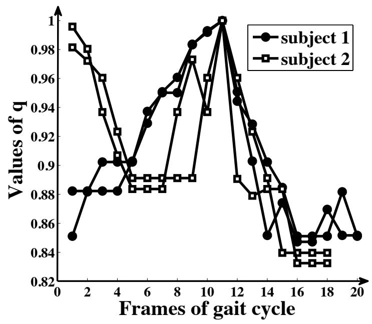


Figure 9: The discrete signals representing the ARPoLK for a gait period of each subject: subject 1 and subject 2.

discrete signals representing ARPoLK for gait periods of lengths (i.e., number of frames of the gait period) $m \in \mathbb{N}$ and $n \in \mathbb{N}$, respectively of a gallery and a probe subject, where θ_{LKgi} and θ_{LKpj} are angles of the leading knee. DTW constructs an $m \times n$ matrix whose each element corresponds to the Euclidean distance $d(\theta_{LKgi}, \theta_{LKpj}) = (\theta_{LKgi} - \theta_{LKpj})^2$.

An $m \times n$ warping path is a sequence $p = (p_1, p_2, ..., p_L)$ with $p_l = (\theta_{LKgm_l}, \theta_{LKpn_l}) \in [1 : m] \times [1 : n]$ for $l \in [1 : L]$ and $\max(m, n) \leq L < (m + n)$, for mapping two sequences a_g and a_p which satisfies the followings: (a) boundary condition: $p_1 = (1, 1)$ and $p_L = (m, n)$; (b) monotonicity condition: $\theta_{LKgm_1} \leq \theta_{LKgm_2} \leq .... \leq \theta_{LKgm_L}$ and $\theta_{LKpn_1} \leq \theta_{LKpn_2} \leq .... \leq \theta_{LKpn_L}$; and (c) step size condition: $p_{l+1} - p_l \in (1, 0), (0, 1), (1, 1)$ for $l \in [1 : L - 1]$. DTW minimises the cost of warping a_g and a_p together to form the third gait signature,

$$\mathrm{d_{DTW}} = \min\left(\frac{(\sum_{l=1}^{L} p_l)^{\frac{1}{2}}}{L}\right). \tag{12}$$

3.5. Module 3: Identify subject

Each of the gallery and probe gait sequences respectively with N_g and N_p frames is partitioned into consecutive subsequences with gallery gait period (G_g) and probe gait period (G_p). The distance metric between the kth probe gait period and a gallery sequence for match score S, where S is either $\mathrm{d_{PWMS}}$, $\mathrm{d_{BDHM}}$ or $\mathrm{d_{DTW}}$, is

$$Dist_S(k) = \min_i(S), \tag{13}$$

where i = 1,2,...,n_g and $n_g = N_g/G_g$ is the number of gallery gait periods in a gallery sequence. The median of the distances

$$D_S = median(Dist_S(1), Dist_S(2), ..., Dist_S(m_p)), \tag{14}$$

is considered as the match score between the probe sequence and gallery sequence to be used in the score-level fusion for subject identification, where $m_p = N_p/G_p$ is the number of probe gait periods in a probe sequence.

Unlike STM-SPP, which uses a rank-based classifier combination rule to combine the classification results by Procrustes shape analysis and EFDs for identifying a subject, STS-DM uses score-level fusion to fuse the match scores obtained by the PWMS, BDHM and DTW. Since score-level fusion combines the match scores obtained by different classifiers, it is more informative than rank-level fusion. Rank-level fusion is also computationally more expensive and suffers from the drawback of a tie in ranking, which requires further processing to get resolved, e.g., STM-SPP uses Hu moments to resolve a tie in ranking. However, score-level fusion requires the inhomogenous scores obtained by different classifiers to be transformed into a common numerical range before being compared using score normalisation technique. The linear score normalisation techniques, e.g., min-max normalisation and z-score normalisation have similar computational complexities, but z-score normalisation is preferred in STS-DM as it is less sensitive to outliers than min-max normalisation. Although non-linear score normalisation techniques based on double sigmoid function and hyperbolic tangent are more robust to outliers, they introduce complexity due to the use of many parameters, and the performance of these techniques are highly dependent on the chosen parameter values.

Hence, to make a trade-off between the performance and computation complexity, STS-DM obtains z-scores of each of the three match score sets using

$$Z_{classifier} = \frac{\mathrm{d}_{classifier} - \mu_{classifier}}{\sigma_{classifier}}, \tag{15}$$

where μ is mean of the set of scores, d is the individual score, σ is standard deviation, and $classifier$ is either PWMS, BDHM or DTW. The three $classifiers$ do not perform equally well as evident from Fig. 11 (see Section 4.1 for CMC curve) which shows that PWMS is the best feature, while BDHM is the worst. A weight-based sum rule of score-level fusion [53] is thus used in STS-DM for improved identification rate where the weights are determined based on the contribution of each component $classifier$ to the final subject identification. The fused score is thus obtained using

$$S_f = \frac{\mathrm{I_{PWMS}} \times Z_{\mathrm{PWMS}} + \mathrm{I_{BDHM}} \times Z_{\mathrm{BDHM}} + \mathrm{I_{DTW}} \times Z_{\mathrm{DTW}}}{\mathrm{I_{PWMS}} + \mathrm{I_{BDHM}} + \mathrm{I_{DTW}}}, \tag{16}$$

where $\mathrm{I_{PWMS}}$, $\mathrm{I_{BDHM}}$ and $\mathrm{I_{DTW}}$ are the weights that respectively correspond to the CCR (see Section 4.1 for CCR) obtained using the match scores $\mathrm{d_{PWMS}}$, $\mathrm{d_{BDHM}}$ and $\mathrm{d_{DTW}}$ for a particular testing condition. The probe subject is identified based on the lowest S_f it measures with the member of a gallery class.

4. Experiments

Since the aim of STS-DM is to demonstrate its combined robustness against most of the challenging factors of gait recognition, it is extensively compared with several related methods that individually address one or more covariate factors. Therefore, to make uniform comparison with several related methods, STS-DM is evaluated using different experimental setup based on the reported available results of those methods on two public data sets: CMU Motion of Body (MoBo) data set and USF HumanID gait challenge data set.

4.1. Experiments on CMU MoBo data set

CMU Motion of Body (MoBo) data set [54] comprises gait sequences of 25 subjects performing four types of walk: slow walk (walking speed: 2.06 mph); fast walk (walking speed: 2.82 mph); walk holding a ball (walking speed: 2.04 mph); and walk on an inclined plane of a treadmill (walking speed: 1.96 mph). Each sequence is of approximately 11 seconds duration and is recorded at 30 frames per second from six different views. The sequences were captured on a same day using six high resolution calibrated cameras evenly distributed around the treadmill.

A closed-set identification guarantees the existence of the subject in the database. We analyse the closed set identification performance of STS-DM on the profile view silhouettes of MoBo data set by taking one subject as the probe sample and train it on all the subjects of the data set including the probe sample. The percentage of correct classification rate is

$$\mathrm{CCR}(\%) = s_c/s_t * 100, \tag{17}$$

where s_c and s_t are respectively the number of correctly identified subjects and the total number of subjects in the data set. The identification is best interpreted by a cumulative match characteristic (CMC) curve which shows CCR at different ranks. Since the smaller the values of the match scores the more similar are the two subjects, the CCR at rank r implies that the probability of correct match is among the lowest r match scores.

We use 3D scatter plots as shown in Fig. 10 to show the distribution of match scores (plotted along the vertical axis) obtained by PWMS, DTW, BDHM and the fused classifier as a result of comparing each of the fast walking probe subjects (plotted along the horizontal right axis) with all 25 slow walking gallery subjects (plotted along the horizontal left axis) of CMU MoBo data set. Note that the ith probe subject along the horizontal right axis corresponds to its matching gallery subject i along the horizontal left axis, where i=1,2,3,...,25. The match scores obtained by comparing one probe with all the gallery subjects are represented by circles of the same sizes in the plots, while different circle sizes are used for different probe subjects. Since ideally a probe subject will result in the lowest match score for its matching gallery subject, very few circles are present in the bottom horizontal planes. A probe subject will generate higher match scores for all the non-matching gallery subjects, which explains why the circles of different sizes are cluttered around the higher horizontal planes of the plots.

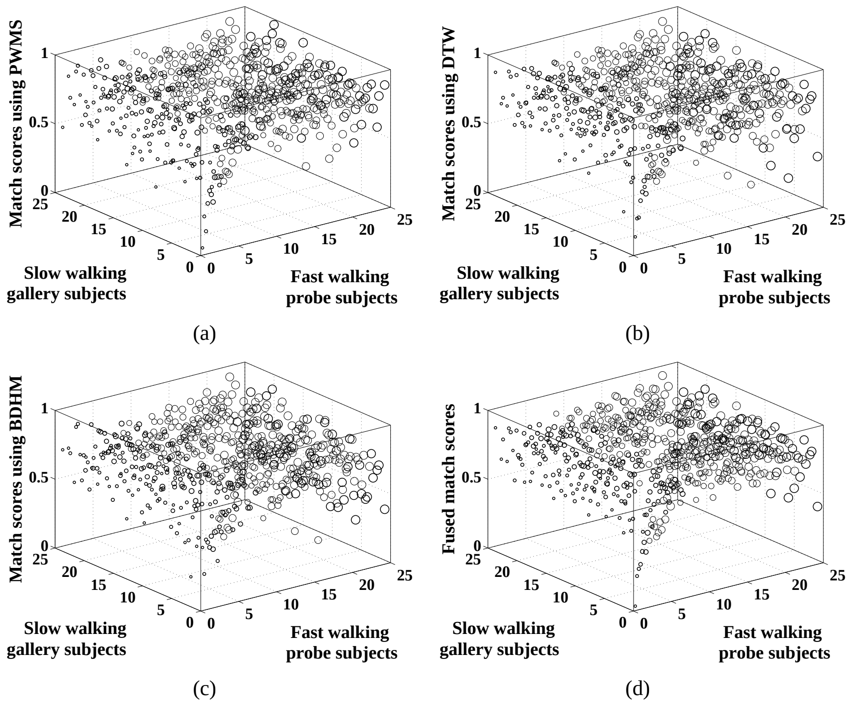
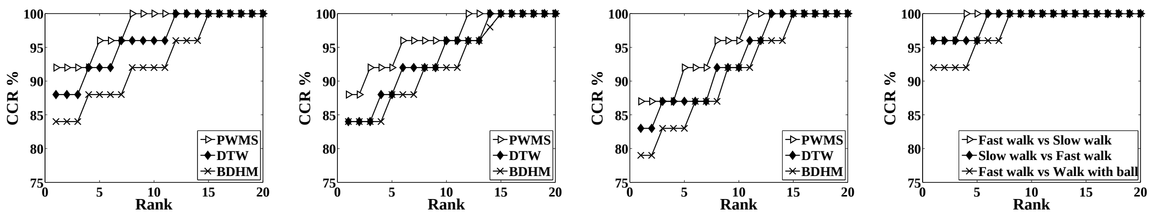


Figure 10: Distribution of match scores obtained by (a) PWMS, (b) DTW, (c) BDHM and (d) fused classifier for fast walk vs slow walk of lateral-view silhouettes of CMU MoBo data set.

Note that the number of probe subjects that results in the lowest match scores for their matching gallery subjects using PWMS, DTW, BDHM and fused classifier are respectively 23, 22, 21 and 24 for fast walk vs slow walk. Hence, the rank-1 CCR for PWMS, DTW, BDHM and fused classifier are respectively 23/25*100, i.e., 92%, 22/25*100, i.e., 88%, 21/25*100, i.e., 84% and 24/25*100, i.e., 96% which are verified in Fig. 11(a), where Fig. 11 shows the CMC curves of CCR obtained using PWMS, DTW and BDHM for three different walking conditions of CMU MoBo data set, namely fast walk vs slow walk, slow walk vs fast walk, and fast walk vs walking with ball. It is clear that the performance of STS-DM is the best for fast walk vs slow walk using individual classifiers as well as the fused classifier. The rank 1 CCR of PWMS, DTW and BDHM are respectively 92%, 88% and 84% for fast walk vs slow walk; 88%, 84% and 84% for slow walk vs fast walk; and 87%, 83% and 79% for fast walk vs walking with ball. Since PWMS outperforms DTW and BDHM, it is shown that the shape of a subject provides better inter-subject discriminative characteristics than its kinematics in the case of very limited across-day gait variations. Fig. 11(d) shows that CCR is significantly improved, i.e., 96%, 96% and 92% respectively for fast walk vs slow walk, slow walk vs fast walk and fast walk vs walking with ball by fusing the results of individual classifiers using weight-based sum rule of score-level fusion.

4.1.1. Comparisons

The performance of STS-DM on the lateral view of silhouettes of the CMU MoBo data set is compared with shape sequence matching (SSM) based method in [17], SSP [16], STM-SPP [23] and SVB frieze [33]. Table 2 shows that the shape based approach in SSM using stance correlation for the subjects walking parallel to the image plane is robust to variations in walking speed, but its performance degrades significantly when the shape of the silhouettes change due to different activities (e.g., fast walk vs walking with ball). Since part-based shape analysis using EFDs and component-based shape analysis using FDs respectively aid STM-SPP and STS-DM to achieve invariance to carrying small items, they significantly outperform SSM and SVB frieze. The superiority of STS-DM over STM-SPP is attributed to the analyses of dynamic motion characteristics of silhouettes using ellipses fitted to various body parts and ARPoLK that enable it to achieve robustness against limited variations in clothing.



(a) (b) (c) (d)

Figure 11: CMC curves of classification rates obtained using PWMS, BDHM and DTW of the lateral-view silhouettes from CMU MoBo data set for (a) fast walk vs slow walk; (b) slow walk vs fast walk; and (c) fast walk vs walking with ball. (d): CMC curves of combined classification rates of (a)-(c) using weight-based sum rule of score-level fusion.

Table 2: Top-rank identification rates (in percentage) on CMU MoBo data set with the rates of SSM from [17], Baseline from [15], CMU from [55], SSP from [16], STM-SPP from [23] and SVB frieze from [33] for the lateral view. Keys: 'G' - Gallery sequence; 'P' - Probe sequence; 'S' - Slow walk; 'F' - Fast walk; 'B' - Walk with ball.

G/P	SSM [17]	Baseline [15]	CMU [55]	SSP [16]	STM-SPP [23]	SVB frieze [33]	STS-DM
S/S	100	92		100	100	100	100
F/F	100	-		100	100	100	100
B/B	92	-		-	100	100	100
S/F	80	72	76	54	94	82	96
F/S	84	-		32	91	80	96
S/B	48	88	92	-	93	77	92
B/S	68	-		-	82	89	92
F/B	48	-		-	84	61	92
B/F	48	-		-	82	73	87

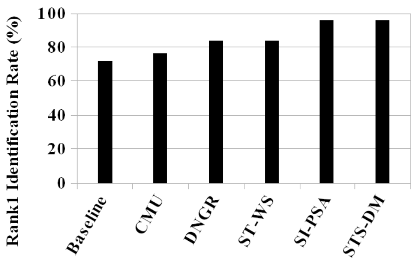


Figure 12: Comparison with related works. Baseline [15], CMU, DNGR [13] and STS-DM are evaluated on CMU MoBo gait data set (experiment 2 of CMU) with walking speed variation of 3.3 km/h and 4.5 km/h, while ST-WS [36] and SI-PSA [35] are evaluated on OU-ISIR treadmill gait data set A [56] with walking speed variation of 3 km/h and 4 km/h between gallery and probe gait sequences.

The method in [16] computes image self-similarity plot (SSP), i.e., correlation of corresponding pairs of images in two gait sequences of a subject. To make uniform comparison with SSP which is robust to segmentation noise, STS-DM also uses split-sample cross validation technique like SSP, where gallery and probe sets correspond to different combination of walking types for each of the twenty-five subjects. Since STS-DM is defined only on profile view of the silhouettes, we consider profile view of two sequences per subject (total 50 sequences) walking at slow pace (2.06 miles/h) and fast pace (2.82 miles/h). Table 2 shows that the performance of SSP for profile view degrades significantly when the probe and gallery samples differ in walking speed. STS-DM outperforms SSP by analysing shape and motion characteristics of ten phases of a gait period and using DTW for ARPoLK analysis so as to overcome the effects of walking speed variations.

To demonstrate robustness against speed variations by comparative experimental analysis with the related speed-invariant methods, STS-DM is evaluated using the experiment 2 defined by the method silhouette-based human identification from body shape and gait (CMU) in [55] as it enables evaluation of a gait recognition method across different speeds. The rank 1 identification rates of STS-DM, speed-invariant method DNGR, CMU [55] and Baseline [15] are respectively 96%, 84%, 76% and 72% (see Fig. 12) for the slow (3.3km/h) vs fast (4.5km/h) walking gait sequences of the profile view silhouettes of CMU MoBo data set, where the rates of DNGR, CMU and Baseline are based on experiment 2 of CMU entitled “Across gaits condition”. Since speed variation in this experiment is almost 1km/h, we compare STS-DM with speed-invariant methods ST-WS [36] and SI-PSA [35], that are evaluated on OU-ISIR treadmill gait data set A [56] with walking speed variation of 3 km/h and 4 km/h between gallery and probe gait sequences. It is clear from Fig. 12 that STS-DM outperforms all other methods, and provides equal rank 1 identification rate as SI-PSA.

4.2. Experiments on USF HumanID gait challenge data set

STS-DM is evaluated on both the small version (452 sequences from 74 subjects, data acquired in May only) and the full version (1870 sequences from 122 subjects, data acquired in May and November) of USF HumanID gait challenge data set [15]. The data set comprises sequences of subjects walking along elliptical paths in an outdoor environment in front of two cameras with the following five covariates: walking surface (grass (G) or concrete (C)); shoe type (A or B); viewpoint (right (R) or left (L)); carrying conditions (carrying a briefcase (BF) or not carrying a briefcase (NB)); and elapsed time between the acquisition of the sequences (May (M) or November (N)). Twelve experiments of increasing difficulty are designed as shown in Table 3 and Table 4 to investigate the robustness of a gait recognition method against the five covariates. The gait sequences are captured at 30 fps, and the spatial resolution of each silhouette is 128×88. The thirty three subjects that are common in the May and November data sets account for time covariate. There are no common sequences between the gallery and the probe sets, and all subjects did not participate in all experiments [4, 15].

4.2.1. Comparisons

Table 3 shows the results on the full version of USF data set in terms of identification rate (P_I) at ranks 1 and 5, to enable a comparison with the state-of-the-art methods, i.e., GTDA-GF [21], GEI [4], RCK-G [19], GFI [22],

Table 3: Identification rates (in percentage) on full version of USF HumanID gait challenge data set using the gallery set (G, A, R, NB, M/N) of 122 subjects. The rates for GTDA-GF, GEI, RCK-G, GFI, CGI, STM-SPP, DNGR and MMFA are from [21], [4], [19], [22], [41], [23], [13] and [14], respectively. Keys for covariates: V - view; H - shoe; S - surface; B - briefcase; T - time; and C - clothes.

Exp[Covariate]	Identification Rate (P_I)%									
	GTDA-GF(GEI)		RCK-G(GFI)		CGI(STM-SPP)		DNGR(MMFA)		STS-DM	
	Rank1	Rank5	Rank1	Rank5	Rank1	Rank5	Rank1	Rank5	Rank1	Rank5
A [V]	91(90)	98(94)	83(89)	96(98)	91(92)	97(96)	85(89)	96(98)	93	97
B [H]	93(91)	99(94)	86(93)	94(94)	93(95)	96(98)	89(94)	94(98)	96	98
C [V, H]	86(81)	97(93)	78(70)	88(93)	78(84)	94(95)	72(80)	89(94)	86	96
D [S]	32(56)	68(78)	39(19)	66(40)	51(72)	77(80)	57(44)	85(76)	70	82
E [S, H]	47(64)	68(81)	34(23)	63(47)	53(68)	77(84)	66(47)	81(76)	69	83
F [S, V]	21(25)	50(56)	20(7)	51(26)	35(29)	56(59)	46(25)	68(57)	39	61
G [S, H, V]	32(36)	56(53)	21(8)	46 (25)	38(40)	58(61)	41(33)	69(60)	37	60
H [B]	95(64)	95(90)	43(78)	66(94)	84(69)	98(92)	83(85)	96(95)	78	95
I [B, H]	90(60)	99(83)	40(67)	68(85)	78(60)	97(84)	79(83)	95(93)	71	89
J [B, V]	68(60)	84(82)	40(48)	65(74)	64(64)	86(85)	52(60)	79(84)	66	83
K [T, H, C]	16(6)	40(27)	16(3)	44(24)	3(20)	27(30)	15(27)	46(48)	27	39
L [S, T, H, C]	19(15)	40(21)	5(9)	22(24)	9(18)	24(27)	24(21)	39(39)	22	28
$\overline{X_I}$	61(58)	78(76)	44(46)	67(64)	62(63)	79(79)	63(60)	82(80)	67	80

CGI [41], STM-SPP [23], DNGR [13] and MMFA [14]. The method using matrix-based marginal Fisher analysis (MMFA) in [14] applies marginal Fisher analysis on GEIs for gait representation to reduce the dimensionality of the feature space and extends marginal Fisher analysis to marginal based analysis for content-based image retrieval. Table 4 shows the results on the full version of USF data set in terms of verification rate (the probability that the method successfully detects the correct match between the probe and gallery sequences, i.e., P_V) at false alarm rates (the probability that the method incorrectly classifies a probe sequence to a nonmatching gallery sequence) 1% and 10% for Baseline, DNGR and STS-DM. Since the number of probe subjects in the gait challenge experiments varies, we compute the weighted average identification rate ($\overline{X_I}$) and the weighted average verification rate ($\overline{X_V}$) [20], i.e.,

$$\overline{X_I} = \frac{\sum_{i=1}^{g} w_i x_i}{\sum_{i=1}^{g} w_i}\ ,\ \overline{X_V} = \frac{\sum_{i=1}^{g} w_i x_v}{\sum_{i=1}^{g} w_i}\ , \tag{18}$$

where g denotes the number of challenge experiments whose values are respectively 12 (i.e., Exp. A-L) and 7 (i.e., Exp. A-G) for the full and small versions of HumanID gait challenge data set. x_i and x_v are respectively the identification and verification rates (in percentage) for the ith challenge experiment, and w_i is the number of probe subjects participating in that experiment.

The identification rates achieved by GEI for the twelve challenge experiments after combining the real and synthetic gait features are presented in Table 3. GTDA-GF reports the identification rates obtained by applying GTDA as a preprocessing step of linear discriminant analysis on the magnitude of the result of convolving a GEI with sum of Gabor functions over scales with direction fixed. The rates of GFI in Table 3 are based on direct matching of gallery and probe sequences using an exemplar GFI for reducing computational complexity. Table 3 shows that STS-DM outperforms all other methods for experiment A with a variation in view, and performs reasonably better than GTDA-GF, GEI, RCK-G, GFI, CGI and STM-SPP for experiments K and L with a variation in clothing. However, STS-DM is outperformed by GTDA-GF, CGI, DNGR and MMFA for the gait challenge experiments H and I. This is because AR-PoLK analysis with a consideration of subject's arm-swing is particularly affected by the briefcase carrying condition as it prevents normal arm-swing and distorts the shape of a silhouette between hip and knee. Since the gait challenge experiments H and I take into account of briefcase covariate, the performance of STS-DM is degraded in these experiments. The aim of STS-DM is to achieve combined invariance to most of the challenging factors of gait recognition

Table 4: Verification rates at a false alarm rate (P_F) of 1% and 10% for Baseline from [15], DNGR from [13], STM-SPP from [23] and STS-DM on full version of USF HumanID gait challenge data set using the gallery set (G, A, R, NB, M/N) of 122 subjects. Keys for covariates: V - view; H - shoe; S - surface; B - briefcase; T - time; and C - clothes.

Exp.	Covariate	Verification Rate (P_V)% at			
		Baseline	DNGR	STM-SPP	STS-DM
		P_F:1(10)%	P_F:1(10)%	P_F:1(10)%	P_F:1(10)%
A	V	82(94)	93(98)	88(100)	94(100)
B	H	87(94)	94(98)	94(100)	97(100)
C	V, H	65(94)	80(94)	86(98)	88(98)
D	S	44(80)	68(96)	80(94)	79(94)
E	S, H	35(76)	62(90)	74(84)	76(84)
F	S, V	20(60)	53(86)	50(82)	66(82)
G	S, H, V	28(55)	43(79)	52(76)	62(76)
H	B	72(91)	91(99)	83(95)	85(95)
I	B, H	67(85)	86(97)	76(93)	76(93)
J	B, V	48(76)	58(92)	65(92)	68(92)
K	T, H, C	6(24)	27(61)	21(58)	29(58)
L	S, T, H, C	6(24)	24(46)	19(52)	25(52)
Weighted average verification rate ($\overline{X_V}$)		51(76)	70(91)	70(89)	75(90)

with low computational complexity, and not to achieve the best identification rates for every gait challenge experiment among the state-of-the-art gait recognition methods. The superiority of STS-DM to other methods in terms of $\overline{X_I}$ and $\overline{X_V}$ is demonstrated in Table 3 and Table 4. Table 3 shows that STS-DM achieves the highest $\overline{X_I}$ at rank 1, followed by DNGR, STM-SPP and CGI, and is only second to DNGR in terms of $\overline{X_I}$ at rank 5. It is clear from Table 4 that in terms of $\overline{X_V}$ STS-DM outperforms other methods at the false alarm rate of 1%.

Table 5 shows the results on the small version of the data set (No-Briefcase data) to enable a comparison with Baseline, silhouette analysis-based gait recognition (CASIA) in [18], gait recognition based on shape estimation (CMU1) in [57], CMU [55], RCK-G [19], GEI [4], ASM [6] and STM-SPP [23]. We present the identification rates at rank 1 of CMU1 obtained by weighted correlation similarity measure, and the identification rates of GEI obtained by fusing real and synthetic gait templates. Table 5 shows that STS-DM achieves the second highest $\overline{X_I}$ following ASM. All methods listed in Table 5 except ASM use the silhouettes provided by the USF HumanID gait challenge data set for uniform comparison. Since these silhouettes are significantly affected by strong shadows under feet (mainly due to the subjects walking on a concrete surface as in the gait challenge experiments D, E, F and G) the methods that directly use the silhouettes provided by the USF HumanID gait challenge data set do not provide satisfactory recognition rates for these experiments. ASM employs hierarchical prediction-based ASM framework with Kalman filter to extract the foreground which is unaffected by shadows, and analyses its model parameters for gait recognition. Hence, the superiority of ASM for the gait challenge experiments D, E, F and G is attributed to the use of shadow-free good quality silhouettes for feature extraction. Also, the feature extraction and classification processes involved in ASM are much more computationally expensive compared to STS-DM. Disregarding the performance of ASM, Table 5 shows that STS-DM outperforms all the methods for all the gait challenge experiments.

The performance of STS-DM for the twelve challenge experiments of the full version of USF HumanID gait challenge data set is measured by identification mode and verification mode, using CMC and Receiver Operating Characteristic (ROC) curves respectively, following [58]. Fig. 13(a) shows that the identification rates of STS-DM range from 22% to 96% at rank 1, and 28% to 98% at rank 5. Fig. 13(b) shows that the verification rates of STS-DM range from 25% to 97% at a false alarm rate of 1%, and 52% to 100% at a false alarm rate of 10%. Table 3 and Fig. 13 show that STS-DM is least affected by variation in shoe types, followed by about 30 degrees change in viewpoint. However, time (i.e., when the data set was generated) has the most impact on the performance of STS-DM, as it

Table 5: Top-rank identification rates (in percentage) on the small version of USF HumanID gait challenge data set (data acquired in May only) using the Gallery Set (G, A, R) of 71 subjects. The rates for Baseline, CASIA, CMU1, CMU, RCK-G, GEI, ASM and STM-SPP are from [15], [18], [57], [57], [19], [4], [6] and [23], respectively. Keys for covariates: V - view; H - shoe; S - surface; B - briefcase; T - time; and C - clothes. Unlike others, identification rates with '$*$' are not based on silhouettes provided by USF HumanID gait challenge data set.

Probe Set	A	B	C	D	E	F	G	
Probe Size	71	41	41	66	42	66	42	
Probe	G,A,L	G,B,R	G,B,L	C,A,R	C,B,R	C,A,L	C,B,L	
Covariate difference	V	H	VH	S	SH	SV	SHV	
Rank-1 Identification Rate (P_I) %								$\overline{X_I}$
Baseline [15]	87	81	54	39	33	29	26	50.62
CASIA [18]	70.42	58.54	51.22	34.33	21.43	27.27	14.29	40.83
CMU1 [57]	85	81	60	23	17	25	21	44.93
CMU [55]	87	81	66	21	19	27	23	46.44
RCK-G [19]	97	89	83	41	34	30	28	57.53
GEI [4]	100	90	85	47	57	32	31	62.83
ASM [6]	97*	95*	91*	92*	86*	85*	78*	89.66
STM-SPP [23]	100	94	89	73	69	40	36	71.74
STS-DM	100	98	91	76	70	47	42	74.99

implies variations in clothing and footwear of the same subject.

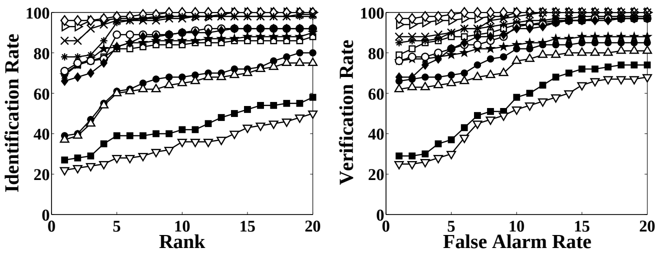
4.3. *Effect of missing frames*

Occlusions in the scene, large shadows under feet and extreme lighting variations can severely distort the extracted contours. If these distorted contours are not part of any of the ten phases of a gait period, they do not affect the classifications using FDs and ellipsoidal fits. If the distortion causes any frame of the ten phases to be missing, its immediate adjacent frame is considered. ARPoLK analysis is not affected if the portion of the contour enclosed in the region between hip and knee remain undistorted. It is also not affected by any missing or discarded frame due to excessive distortions resulting in different lengths of gait sequences. This is because the use of DTW in ARPoLK analysis enables detection of similarity between two sequences of varying lengths. Hence, STS-DM is robust to severely distorted and missing frames.

To support the claim that STS-DM is robust to missing frames by experimental results, we create probe gait sequences of shorter lengths from CMU MoBo data set by discarding frames at a specified interval in order to stimulate a situation where probe frames are missing. In Fig. 14, the rank-1 CCR is plotted along vertical axis, while the horizontal axis shows the intervals of missing frames in terms of number of frames, i.e., 6 at the horizontal axis denotes that every 7th frame is missing from the entire probe sequence. Fig. 14(a)-(c) respectively show the effect of missing frames on rank 1 CCR of STS-DM using fused classifier, PWMS, DTW and BDHM for three testing conditions of CMU MoBo data set, namely fast walk vs slow walk, slow walk vs fast walk and fast walk vs walking with ball. It is evident from the three plots that the rank 1 CCR of STS-DM is not affected for at least every 8th frame is missing from the probe sequence for any of the component classifiers and the fused classifier for three testing conditions. Note that DTW is less robust against missing frames than PWMS and BDHM.

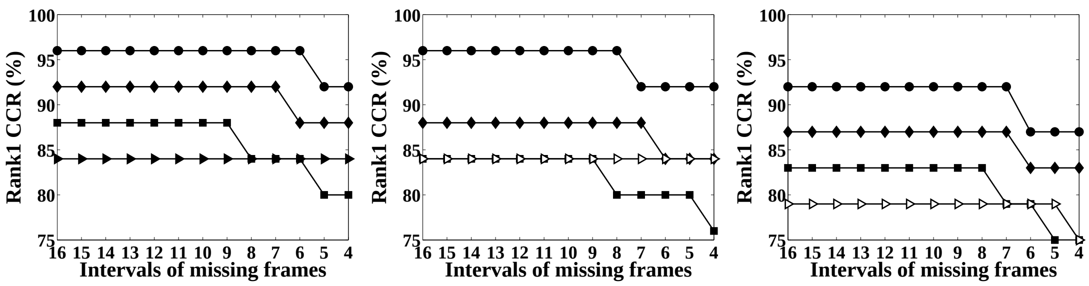
4.4. *Computational complexity*

The computational time of STS-DM is measured using the computer system clock and OpenCV 2.1 in Microsoft Visual Studio 2008 Express Edition environment on an Intel (R) Core (TM) i7 processor working at 2.93 GHz with 4 GB RAM running Windows 7 operating system. For the silhouettes of the full version of USF HumanID gait challenge data set, the processing time for comparing all ten Rf-ROIs one at a time with the tr-ROIs for extracting ten phases of a gait period based on the lowest S_{score} using weighted Krawtchouk moments is 5 sec. The processing time to compute



(a) (b)

Figure 13: Performance on twelve challenge experiments of USF data set. (a) Identification mode (CMC) and (b) Verification mode (ROC). Keys: '▹'- Exp. A (Probe: G, A, L, NB, M/N); '◇'- Exp. B (Probe: G, B, R, NB, M/N); '×'- Exp. C (Probe: G, B, L, NB, M/N); '□'- Exp. D (Probe: C, A, R, NB, M/N); '★'- Exp. E (Probe: C, B, R, NB, M/N); '●'- Exp. F (Probe: C, A, L, NB, M/N); '△'- Exp. G (Probe: C, B, L, NB, M/N); '∗'- Exp. H (Probe: G, A, R, BF, M/N); '○'- Exp. I (Probe: G, B, R, BF, M/N); '◆'- Exp. J (Probe: G, A, L, BF, M/N); '■'- Exp. K (Probe: G, A/B, R, NB, N); and '▿'- Exp. L (Probe: C, A/B, R, NB, N).



(a) (b) (c)

Figure 14: Effects of missing frames on performance using PWMS, DTW, BDHM and fused classifier on CMU MoBo data set: (a) fast walk vs slow walk; (b) slow walk vs fast walk; and (c) fast walk vs walking with ball. Keys: '◆'- PWMS; '■'- DTW; '▹'- BDHM; '●'- Fused Classifier.

d_{PWMS}, d_{BDHM} and d_{DTW} between a probe and a gallery subject is approximately 0.77 fps. Since the Baseline method is characterised by unlimited spatio-temporal correlation of silhouettes, it has very high computational complexity. The hierarchical prediction-based ASM framework with Kalman filter used in ASM to analyse static and dynamic gait characteristics is also computationally very expensive. Unlike most gait recognition methods which process sequences of silhouettes, the real-time method in [59] analyses the set of the largest rectangles fitted onto silhouettes over a gait period spanning up to 25 frames to reduce the computational complexity. STS-DM further reduces this by analysing the shape of contour instead of silhouette at the ten phases of a gait period in computing d_{PWMS} and d_{BDHM}. Since ARPoLK analysis over a gait period uses a1D signal, it does not significantly increase computational complexity. It takes about 5 sec/gait-period to obtain the ten phases by comparing small subregions of an image, i.e., Rf-ROIs with Tr-ROIs, thus reducing time and space complexity.

The use of Cooley-Tukey Fast Fourier transform algorithm [60, 46] reduces the quadratic time complexity of discrete Fourier transform and its inverse to O(T log T). d_{PWMS} is obtained by analysing the low-frequency FDs of the contour points to reduce the computational complexity. DTW has a quadratic time and space complexity, i.e., O(*mn*), where *m* and *n* denote the length of the sequences being compared. However, it is used to compare short sequences, as the number of constituent frames of a gait period usually range between 18-35. Since STS-DM uses a simplified feature space, it does not require any dimensionality reduction technique like principal component analysis and multiple discriminant analysis as in [4, 18]. Since sum rule of score-level fusion and z-score normalisation only require subtraction by mean, division by standard deviation and summation of normalised scores, it has less computational complexity, i.e., O(N), than the rank-based classifier combination rule which requires sorting score of computational complexity O(N log N), where N is the gallery size, followed by post processing to resolve tie in ranking.

5. Conclusion

Unlike existing systems which only address one or more challenging factors of gait recognition, STS-DM combines spatio-temporal shape and dynamic motion characteristics of silhouette contours to identify a human subject in the presence of most challenging factors of gait recognition with low computational complexity. It analyses the shape of a subject by FDs at ten phases of a gait period and introduces a component-based FD analysis to achieve robustness against shape distortion due to all common types of small carrying conditions with folded hands, at the subject's back and in upright position. ARPoLK analysis with consideration of the integral relationship between the motion of limbs and arm-swing enables STS-DM to achieve robustness against gait variations over different days, e.g., limited clothing variations, hair style, shadows under feet and missing body parts. The similarity between the ARPoLK of two subjects is measured using DTW to achieve invariance to walking speed. STS-DM uses BDHM to analyse the full-body shape and motion characteristics by fitting ellipses to five different parts of the human body which is invariant to boundary shape distortions due to segmentation imperfections and missing body parts. The match scores obtained by analysing the local and global gait characteristics using the three feature extractors are combined using weight-based sum rule of score-level fusion for subject identification.

STS-DM analyses the shape of contours, hence it is insensitive to colour and texture of subject's clothing. The feature space of STS-DM does not require any dimensionality reduction. The excellent identification rates in the presence of various challenging factors demonstrate the efficacy of STS-DM. Being a contour based method, STS-DM has a low computational complexity, but it is sensitive to segmentation imperfections, and its performance largely depends on preprocessing. Also, STS-DM is designed for lateral views of gait sequences, thus future developments are required to enable STS-DM to address unconstrained human movements especially in cluttered scenes.

Acknowledgments

The authors would like to thank Warwick Postgraduate Research Scholarship for providing the funds for this research.

[1] W. Hu, T. Tan, L. Wang, S. Maybank, A survey on visual surveillance of object motion and behaviours, IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., 34 (3) (2004) 334-352.

[2] M. S. Nixon, J. N. Carter, Automatic Recognition by Gait, in: Proc. IEEE, vol. 94, no. 11, 2006, pp. 2013-2024.

[3] M. S. Nixon, T. N. Tan, R. Chellappa, Human identification based on gait, Springer Science+Business Media, New York, USA, 2006.

[4] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE Trans. Pattern Anal. Mach. Intell., 28 (2) (2006) 316-322.
[5] L. Wang, H. Ning, T. Tan, Fusion of static and dynamic body biometrics for gait recognition, IEEE Trans. Circuits Syst. Video Technol., 14 (2) (2004) 149-158.
[6] D. Kim, D. Kim, J. Paik, Gait recognition using active shape model and motion prediction, IET Comput. Vis., 4 (1) (2010) 25-36.
[7] F. Tafazzoli, R. Safabakhsh, Model-based human gait recognition using leg and arm movements, Eng. Appl. Artif. Intell., 23 (2010) 1237-1246.
[8] S. L. Dockstader, M. J. Berg, A. M. Tekalp, Stochastic kinematic modeling and feature extraction for gait analysis, IEEE Trans. Image Process., 12 (8) (2003) 962-976.
[9] Z. Xue, D. Ming, W. Song, B. Wan, S. Jin, Infrared gait recognition based on wavelet transform and support vector machine, Pattern Recognit., 43 (2010) 2904-2910.
[10] L. Lee, W. E. L. Grimson, Gait analysis for recognition and classification, in: Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2002, pp. 148-155.
[11] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, A full-body layered deformable model for automatic model-based gait recognition, EURASIP J. Adv. Sig. Pr., (2008) 1-13.
[12] J. Gu, X. Ding, S. Wang, Y. Wu, Action and gait recognition from recovered 3-D human joints, IEEE Trans. Syst., Man, Cybern. B, Cybern., 40 (4) (2010) 1021-1033.
[13] Z. Liu, S. Sarkar, Improved gait recognition by gait dynamics normalization, IEEE Trans. Pattern Anal. Mach. Intell., 28 (6) (2006) 863-876.
[14] D. Xu, S. Yan, D. Tao, S. Lin, H. Zhang, Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval, IEEE Trans. Image Process., 16 (11) (2007) 2811-2821.
[15] S. Sarkar, P. J. Philips, Z. Liu, I. Vega, P. Grother, K. Bowyer, The HumanID gait challenge problem: data sets, performance, and analysis, IEEE Trans. Pattern Anal. Mach. Intell., 27 (2) (2005) 162-177.
[16] C. BenAbdelkader, R. G. Cutler, L. S. Davis, Gait recognition using image self-similarity, EURASIP J. Adv. Sig. Pr., 4 (2004) 572-585.
[17] A. Veraraghavan, A. K. Roy-Chowdhury, R. Chellappa, Matching shape sequences in video with applications in human movement analysis, IEEE Trans. Pattern Anal. Mach. Intell., 27 (12) (2005) 1896-1909.
[18] L. Wang, T. Tan, H. Ning, W. Hu, Silhouette analysis-based gait recognition for human identification, IEEE Trans. Pattern Anal. Mach. Intell., 25 (12) (2003) 1505-1518.
[19] D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, K. Moustakas, Gait recognition using compact feature extraction transforms and depth information, IEEE Trans. Inf. Forensics Security, 2 (3) (2007) 623-630.
[20] D. Xu, Y. Huang, Z. Zeng, X. Xu, Human gait recognition using patch distribution feature and locality-constrained group sparse representation, IEEE Trans. Image Process., 21 (1) (2012) 316-326.
[21] D. Tao, X. Li, X. Wu, S. J. Maybank, General tensor discriminant analysis and Gabor features for gait recognition, IEEE Trans. Pattern Anal. Mach. Intell., 29 (10) (2007) 1700-1715.
[22] T. H. W. Lam, K. H. Cheung, J. N. K. Liu, Gait flow image: A silhouette-based gait representation for human identification, Pattern Recognit., 44 (2011) 973-987.
[23] S. D. Choudhury, T. Tjahjadi, Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors, Pattern Recognit., 45 (2012) 3414-3426.
[24] S. D. Mowbray, M. S. Nixon, Automatic gait recognition via Fourier descriptors of deformable objects, in: Proc. Audio- and Video-based Biometric Person Authentication (AVBPA), Guildford, UK, 2003, pp. 556-573.
[25] Z. Ling, C. Zhao, Q. Pan, Y. Wang, Y. Cheng, Analyzing human movements from silhouettes via Fourier descriptor, in: Proc. IEEE Int. Conf. Automation and Logistics (ICAL), Jinan, China, 2007, pp. 231-236.
[26] I. Venkat, P. D. Wilde, Robust gait recognition by learning and exploiting sub-gait characteristics, Int J. Comput. Vis., 91 (2011), 723, doi 10.1007/s11263-010-0362-6.
[27] M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon, Self-calibrating view-invariant gait biometrics, IEEE Trans. Syst., Man, Cybern. B, Cybern., 40 (4) (2010) 997-1008.
[28] W. Kusakunniran, Q. Wu, J. Zhang, Y. Ma, H. Li, A New View-Invariant Feature for Cross-View Gait Recognition, IEEE Trans. Inf. Forensics Security, doi:10.1109/TIFS.2013.2252342.
[29] S. Zheng, J. Zhang, K. Huang, R. He, T. Tan, Robust view transformation model for gait recognition, in: Proc. IEEE Int. Conf. Image Process., 2011, pp. 2073-2076.
[30] N. Liu, J. Lu, Y. Tan, Joint subspace learning for view-invariant gait recognition, IEEE Signal Process. Lett., 18 (7) (2011) 431- 434.
[31] K. Bashir, T. Xiang, S. Gong, Cross-view gait recognition using correlation strength, in: Proc. British Machine Vis. Conf., London, UK, 2010, pp. 1-11.
[32] G. Zhao, G. Liu, H. Li, and M. Pietikainen, 3D gait recognition using multiple cameras, in: Proc. IEEE Int. Conf. Autom. Face Gesture Recog., Southampton, U.K., 2006, pp. 529 534.
[33] S. Lee, Y. Liu, R. Collins, Shape variation-based frieze pattern for robust gait recognition, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Minneapolis, MN, 2007, pp. 1-8.
[34] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, Gait recognition across various walking speeds using higher order shape configuration based on a differential composition model, IEEE Trans. Syst., Man, Cybern. B, Cybern., 42 (6) (2012) 1654-1668.
[35] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, Speed-invariant gait recognition based on Procrustes shape analysis using higher-order shape configuration, in: Proc. Int. Conf. Image Process., Sydney, Australia, 2011, pp. 545-548.
[36] A. Tsuji, Y. Makihara, Y. Yagi, Silhouette transformation based on walking speed for gait identification, in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., San Francisco, CA, USA, 2010, pp. 717-722.
[37] M. R. Aqmar, K. Shinoda, S. Furui, Robust gait recognition against speed variation, in: Proc. Int. Conf. Pattern Recog., Tokyo, Japan, 2010, pp. 2190-2193.
[38] Y. Ran, Q. Zheng, R. Chellappa, T. M. Strat, Applications of a simple characterization of human gait in surveillance, IEEE Trans. Syst., Man, Cybern. B, Cybern., 40 (4) (2010) 1009-1020.

[39] D. S. Matovski, M. S. Nixon, S. Mahmoodi, J. N. Carter, The effect of time on gait recognition performance, IEEE Trans. Inf. Forensics Security, 7 (2) (2012) 543-552.
[40] M. A. Hossain, Y. Makihara, J. Wang, Y. Yagi, Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control, Pattern Recognit., 43 (2010) 2281-2291.
[41] C. Wang, J. Zhang, L. Wang, J. Pu, X. Yuan, Human identification using temporal information preserving gait template, IEEE Trans. Pattern Anal. Mach. Intell., 34 (11) (2012) 2164-2176.
[42] K. Bashir, T. Xiang, S. Gong, Gait recognition without subject cooperation, Pattern Recognit. Lett., 31 (2010) 2052-2060.
[43] J. Zhang, J. Pu, C. Chen, R. Fleischer, Low-resolution gait recognition, IEEE Trans. Syst., Man, Cybern. B, Cybern., 40 (4) (2010) 986-996.
[44] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, MPCA: Multilinear principal component analysis of tensor objects, IEEE Trans. Neural Netw., 19 (1) (2008) 18-39.
[45] Z. Liu, S. Sarkar, Effect of silhouette quality on hard problems in gait recognition, IEEE Trans. Syst., Man, Cybern. B, Cybern., 35 (2) (2005) 170-183.
[46] G. Bradski, A. Kaehler, Learning OpenCV Computer Vision with the OpenCV Library, O'Reilly Media, Sebastopol, 2008.
[47] D.A. Winter, Biomechanics and motor control of human movement, 3rd ed., John Wiley & Sons, New Jersey, 2004.
[48] P. Yap, R. Paramesran, S. Ong, Image Analysis by Krawtchouk Moments, IEEE Trans. Image Process., 12 (11) (2003) 1367-1377.
[49] D. Cunado, M. S. Nixon, J. N. Carter, Automatic extraction and description of human gait models for recognition purposes, Comput. Vis. Image Und., 90 (1) (2003) 1-41.
[50] S. Yu, D. Tan, T. Tan, A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition, in: Proc. Int. Conf. Pattern Recog., Hong Kong, China, 2006, pp. 441- 444.
[51] C. Wang, J. Zhang, J. Pu, X. Yuan, L. Wang, Chrono-Gait Image: A novel temporal template for gait recognition, in: Proc. Eur. Conf. Comput. Vis., Part 1, LNCS 6311, Heraklion, Greece, 2010, pp. 257-270.
[52] M. S. Nixon, A. S. Aguado, Feature extraction and image processing, 2nd ed., Elsevier, London, 2006.
[53] D. Maltoni, D. Maio, A. K. Jain, S. Prabhakar, Handbook of fingerprint recognition, 2nd ed., Springer Science+Business Media, 2009.
[54] R. Gross, J. Shi, The CMU Motion of Body (MoBo) Database, Tech. report CMU-RI-TR-01-18, Robotics Institute, CMU, 2001.
[55] R. Collins, R. Gross, J. Shi, Silhouette-based human identification from body shape and gait, in: Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2002, pp. 351-356.
[56] Y. Makihara, H. Mannami, A. Tsuji, M. A. Hossain, K. Sugiura, A. Mori, Y. Yagi, The OU-ISIR gait database comprising the treadmill data set, IPSJ Trans. Comput. Vis. Appl., Tech. note, vol. 4, pp. 53-62, 2012.
[57] D. Tolliver, R. T. Collins, Gait shape estimation for identification, in: Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA), Guildford, U.K., 2003, pp. 734-742.
[58] J. Phillips, H. Moon, S. Rizvi, P. Rause, The FERET evaluation methodology for face recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell., 22 (10) (2000) 1090-1104.
[59] O. Barnich, M. V. Droogenbroeck, Frontal-view gait recognition by intra- and inter-frame rectangle size distribution, Pattern Recognit. Lett., 30 (9) (2009) 893-901.
[60] J. W. Cooley, O. W. Tukey, An algorithm for the machine calculation of complex Fourier series, Math. Comput., 19 (90) (1965) 297-301.

Vitae

Sruti Das Choudhury received B.Tech. in Information Technology from West Bengal University of Technology, India and M.Tech. in Computer Science and Application from Calcutta University, India, in 2005 and 2009, respectively. She is currently pursuing Ph.D. in Engineering at the University of Warwick, U.K. Her research interest is gait recognition.

Tardi Tjahjadi received B.Sc. in Mechanical Engineering from University College London in 1980, and M.Sc. in Management Sciences in 1981 and Ph.D. in Total Technology in 1984 from UMIST, U.K. He has been an associate professor at Warwick University since 2000. His research interests include image processing and computer vision.